DOCUMENT RESUME

ED 347 169 TM 018 523

AUTHOR Wilkinson, Rebecca L.
TITLE Statistical Significance Testing: Alternatives and Considerations.
PUB DATE Jan 92
NOTE 28p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (Houston, TX, January 31-February 2, 1992).
PUB TYPE Viewpoints (Opinion/Position Papers, Essays, etc.) (120) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Data Interpretation; Effect Size; *Reliability; Researchers; Research Methodology; *Research Problems; *Statistical Significance
IDENTIFIERS Confidence Intervals (Statistics); Power (Statistics); Research Replication

ABSTRACT

Problems inherent in relying solely on statistical significance testing as a means of data interpretation are reviewed. The biggest problem with statistical significance testing is that researchers have used the results of this testing to ascribe importance or meaning to their studies where such meaning often does not exist. Often researchers mistake statistically significant results for important effects. Statistical procedures are too often used as substitutes to thought, rather than as aids to researcher thinking. Alternatives to statistical significance testing that are explored are effect size, statistical power, and confidence intervals. Other considerations for further data analysis that are explored are: (1) measurement reliability; (2) data exploration; and (3) the replicability of research results. It is suggested that statistical significance testing be used only as a guide in interpreting one's results. Two tables present illustrative information, and there is a 22-item list of references. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

REBECCA L. WILKINSON

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED347169

Statistical Significance Testing:

Alternatives and Considerations

Rebecca L. Wilkinson

Texas A&M University 77843-4225

Paper presented at the annual meeting of the Southwest Educational Research Association, Houston, TX, January 31, 1992.

TM018523

BEST COPY AVAILABLE

Abstract

Over the last decade, statistical significance testing had faced much criticism. The biggest problem with statistical significance testing is that researchers have used the results of this testing to ascribe importance or meaning to their studies where such meaning often does not exist. Statistical significance testing is indeed a valuable tool, but it is not the arbiter of quality in research. It seems that many, if not most, researchers have misinterpreted the meaning of results of statistical significance testing. Often, researchers mistake statistically significant results for important effects. It is suggested that statistical significance testing be used only as a guide in interpreting one's results. The present paper explores the nature of the problems inherent in relying solely on statistical significance testing as a means of data interpretation. Alternatives to statistical significance testing are explored as well as considerations for further data analysis.

Statistical Significance Testing:
Alternatives and Considerations

Statistical significance testing has faced much criticism, especially during the last decade (e.g., Carver, 1978; Meehl, 1978; Thompson, 1987). Some researchers, such as Cronbach (1975) and Carver (1978), even suggest doing away with statistical significance testing altogether. One of the more emphatic views of statistical significance testing is given by Meehl (1978), responding to Sir Ronald Fisher's contributing effort regarding the null hypothesis:

> I suggest to you that Sir Ronald has befuddled us, mesmerized us, and led us down the primrose path. I believe that the almost universal reliance on merely refuting the null hypothesis is the standard method for corroborating substantive theories in the soft (i.e., social science) areas is a terrible mistake, is basically unsound, poor scientific strategy, and one of the worst things that ever happened in the history of psychology . . . I am not making some nit-picking statisticians correction. I am saying that the whole business is so

> radically defective as to be scientifically almost pointless. (pp. 817; 823)

It seems that the biggest problem with statistical significance testing is that researchers have used the results of this testing to ascribe importance or meaning to their studies where such meaning often does not exist. For example, Hays (1988) notes that statistical significance is nothing more than a statement regarding the likelihood of some observed result, i.e., it does not guarantee the importance or value of a finding. He also states that statistical significance testing is a valuable tool, but is not the arbiter of quality. Carver (1978) has this to say about significance testing in research:

> Statistical significance simply means statistical rareness. Results are 'significant' from a statistical point of view because they occur very rarely in random sampling under the conditions of the null hypothesis . . . By itself, statistical significance means little or nothing. (p. 383)

In general, statistical significance testing only allows the researcher to make a yes/no decision

concerning rejection or non-rejection of the null hypothesis (Carver, 1978; Keaster, 1988; Kupfersmid, 1988; Rozenboom, 1970; Shaver, 1980; Thompson, 1988b). All too often, researchers have viewed significance testing as the main goal. As Yates (1951) notes, "Results are significant or not significant and that is the end of it" (p. 33). Huberty (1987) cautions against the use of statistical testing to arrive at a definitive answer, however, and suggests that statistical significance testing be used only as a guide in interpreting one's results. Rosnow and Rosenthal (1989) also warn of the dangers of emphasizing dichotomous, yes/no thinking when it comes to statistical significance testing.

The present paper explores the nature of the problems inherent in relying solely on statistical significance testing as a means of data interpretation. Alternatives to statistical significance testing are explored as well as considerations for further data analysis.

Factors Affecting the Use of Statistical Significance Testing

Misinterpretation of Statistical Significance Testing

It appears that many, if not most, researchers have misinterpreted the meaning of results of statistical significance testing. Shaver (1980) and Thompson (1987, 1989, 1992) note that, often, researchers mistake statistically significant results for important effects. Thompson (1988b) points out that one of the most common mistakes made by graduate researchers is to overvalue significance testing. Just because a result is statistically significant does not mean it has great value in and of itself, i.e., that the results are of any importance.

Of course, it is easy to see how this misinterpretation of results could happen. Craig, Eison, and Metze (1976) note that in a study of selected psychology journals, 50 out of 51 reported the use of statistical significance as an index of the meaningfulness or value of the study. Furthermore, the authors discovered that a finding with a $p<.001$ was given more value than a study with $p<.05$, implying that the value of a research study is associated with lower probability levels. Most importantly, sample sizes

were ignored by most of these authors when they interpreted their results.

Kupfersmid (1988) and Thompson (1987) also noted that journal editors and reviewers are biased toward publishing almost exclusively those research results which reach statistical significance, thereby, reinforcing the myth of the all importance of statistical significance in research. Probability and so-called "significance" are too often seen as end products of analysis and not the first step in the explanation of concomitant variation in the data.

In evaluating the meaningfulness or value of statistical outcomes, Craig, Eison, and Metze (1976) suggest that this meaningfulness not be based just on the associated probability of the statistical test, but on outcomes of greatest interest for usefulness, theory, or even societal implications. This raises the issue of statistical as opposed to substantive significance. This implies that a result could be statistically significant, but have no relative importance for society.

Sample Size and Significance Testing

Much is currently being written about the effect of sample size on significance testing (e.g.,

Kupfersmid, 1988; Schneider & Darcy, 1984; Thompson, 1987, 1989). Hays (1988) notes that any research study can produce statistically significant results given a large enough sample size.

According to Schneider and Darcy (1984), seven factors can be said to determine the outcome of a significance test. These factors are:

(1) actual strength of impact;

(2) number of cases used in the study;

(3) variation among cases on relevant variables;

(4) the complexity of the analysis (degrees of freedom);

(5) the appropriateness of the statistical measures and tests used;

(6) the hypotheses tested; and

(7) the significance level chosen.

Of these factors, the number of cases used in the study has a large impact on the outcome of a significance test. Thompson (1988a, 1988b) concurs, noting that "Sample size is the primary influence on significance" (p. 5).

Kupfersmid (1988) notes that it is rare that the correlation or difference between groups is ever exactly zero. Therefore, the null hypothesis or

hypothesis of no difference is rarely ever retained. This means that with a large enough sample size, virtually any null hypothesis can be rejected. According to Davidson and Giroir (1987):

> The decision to reject or not reject the null hypothesis must be interpreted in the context of the sample size. A significant result that would no longer be significant with a few fewer subjects is troublesome; a nonsignificant result that would be significant with a few more subjects certainly may be noteworthy. (p. 4)

The real question, then, is not whether test results are statistically significant, but how large the sample size must be to obtain statistical significance (Kupfersmid, 1988).

Craig, Eison, and Metze (1976) also note that probability of a statistical test is greatly influenced by the sample size, other factors being held constant, and is not a reflection of the degree of relationship said to exist between the dependent and independent variables. With a large enough sample size, one can obtain statistically significant results even though

the association between the dependent variable(s) and the independent variable(s) is minimal.

Alternatives To Statistical Significance Testing

Due to the controversy surrounding statistical significance testing and the misinterpretations of these tests, some alternative approaches have been offered. Although significance testing does have some usefulness, it should not be used as the only source of interpretation of research results. Shaver (1980) proposes that researchers view statistical significance testing as only one of many criteria used in judging results. Glass (1979) emphasizes thorough examination and analysis of results in as many ways as possible to make sense of one's findings.

Effect Size

An important part of any research is effect size, defined by Kupfersmid (1988) as the degree of relationship or correlation between variables. Shaver (1980) notes that effect size is an important issue; however, educational research paradigms have not conditioned educational researchers to think about the magnitude that an effect size should reach to be considered of importance. Instead, our present conditioning is to set an alpha level, giving

statistical significance testing the most importance, while treating educational or practical significance (and, hence, effect size) as secondary (Craig, Eison, & Metze, 1976; Shaver, 1980).

Unlike statistical significance testing, effect size is not so strongly influenced by the size of the research sample. As Davidson and Giroir (1987) note, "The effect size . . . is not influenced by sample size and therefore results are less ambiguous when effect sizes are reported and interpreted" (p. 6). The authors go on to say that as sample size increases, effect size still remains constant. Therefore, even when research is found to be statistically significant (the null hypothesis is rejected), the effect size may be small. That is, even though test results were statistically significant, the amount of variance explained could be small.

In contrast, results of one's research may not be statistically significant (the null hypothesis is not rejected), but effect size could be quite large. This means that even though results were not statistically significant, a larger amount of variance would be explained. Therefore, effect size is an important

indicator of explained variance not influenced directly by the size of the sample.

For example, problems in evaluating research can arise when sample size is small, and results are not statistically significant. The effect size could be of a magnitude such that the findings are of considerable merit, but the researcher may wrongly conclude that there is no effect due to a lack of statistical significance.

It is rare that researchers look for the effect size, and it is often erroneously concluded that a study did not detect any true effect if statistical significance was not reached. The smaller the sample size the less likely it is that an effect size will be interpreted. Too often, research goes no further with the investigation of statistical significance because of a nonstatistically significant result associated with a small sample even with the accompaniment of a large effect size.

Tables 1 and 2 illustrate these points, using fixed effect sizes and increasing sample sizes. Notice that increasing the sample size leads to a smaller critical F ratio as degrees of freedom error increases. Also, as the degrees of freedom error increases, mean

square error becomes smaller; therefore, the calculated F ratio becomes larger.

In reviewing Table 1, the difference between not rejecting and rejecting the null hypothesis is based on the difference between 53 and 63 subjects, even though the effect size remains fixed at a moderate level of 9.8 percent. In Table 2, the difference is in using a sample of 13 and 23 subjects, even though the effect size in this case is set at 33.6 percent. In both cases, with a statistically significant result being a product of inclusion of about 10 more subjects, one can see that a person with only superficial understanding of the concept of statistical significance could easily misinterpret research results if effect size and sample size were not taken into account.

Insert Tables 1 and 2 about here

The importance of effect size is also noted by Thompson (1987) who states that all analyses are really testing association and that we are really always interested in explaining variance. Hays (1988) states that significance level alone tells us close to nothing about the strength of association between variables.

Craig, Eison, and Metze (1976) argue for the use of some index of the association of variables that would evaluate the relationship strength regardless of which test of significance is used. Thompson (1988b) notes that effect size steers our interpretation toward what is really important in research, bringing up the issue of noteworthiness of results.

Power

Another important complement to significance testing is power analysis. Power analysis estimates the probability of a Type II error (did not reject the null hypothesis when it was true) (Thompson, 1987). Cohen (1988) defines power as the probability that a statistical test will yield results that are statistically significant.

Cohen (1988) and Shaver (1980) note that this approach is rarely used by researchers due to the difficulty inherent in specifying a priori desired or expected effects. The use of a power analysis includes dealing with alpha, effect size, N, and Beta. This type of analysis is still a function of statistical significance. One can manipulate the alpha level, or manipulate N to increase power. These manipulations, while increasing knowledge about all the elements that

are required for decision-making, can still lead to meaningless statistical results. Thompson (1987) notes that if power is high, nonstatistically significant results still contribute to the knowledge base.

Confidence Intervals

A third alternative is the use of confidence intervals (Huberty, 1979; Shaver, 1980). Sample data are used to find the range within which one predicts the population value to exist. For example, suppose one wants to estimate a population mean, μ, with 95 percent confidence. The level of confidence is actually 1 minus the alpha level or the level of significance. If the alpha level is set at .05, then the level of confidence would be 1-.05, or .95. A general formula for a confidence interval is:

CI=(statistic) +/- (critical value) (standard error)

The statistic in this example would be the sample mean, X. If the population variance is known, the standard error is $\sigma_{\bar{X}}$, and can be computed as $\sigma_{\bar{X}}=\sigma/\sqrt{n}$. A normal distribution table in a statistics text (e.g., Hinkel, Wiersma, & Jurs, 1988) can be used to find critical values associated with the desired level of confidence. The critical value represents the points within which

the range of values associated with the confidence interval is contained.

To illustrate with an example, suppose the sample mean, X, equals 30; the standard deviation equals 2; the sample size (n) equals 40; and the calculated standard error, $\sigma_{\bar{x}}$, equals $2/\sqrt{40} = 2/6.324 = .32$. The formula for a 95 percent confidence interval using this information is:

$$CI = X \pm (1.96)(\sigma_{\bar{x}})$$
$$= 30 \pm (1.96)(.32)$$
$$= 30 \pm .63$$
$$= 29.37, 30.63$$

This means that one can conclude with 95 percent confidence that the interval from 29.37 to 30.63 contains the population parameter, or μ.

Rozenboom (1970) notes that the use of confidence intervals might steer researchers away from the inevitable yes/no dichotomy that accompanies significance testing and decision making. Interval estimation may not be appropriate for some situations, however, such as with some nonparametric tests and some multivariate analyses (Huberty, 1979).

Other Considerations in Analyzing Data

Measurement Reliability

Another area to explore when analyzing one's data is measurement reliability. Theoretically, statistically significant results are only possible with reliably measured variables. It is not a test instrument itself that is reliable, but the scores or data that are reliable (Thompson, 1987). It is up to the researcher to establish proof that the data that are collected from all occasions are reliable. Just because one's measure is constant across many studies is not total proof that data from this measure will be reliable. Thompson (1987, 1992) also notes that one must take into consideration and evaluate all factors that yield any given result in order to interpret the results in a meaningful way.

Data Exploration

Another issue is data exploration (Shaver, 1980). Researchers must be encouraged to thoroughly explore their data rather than just relying on significance testing. Researchers tend to focus on purely descriptive statistics, especially the mean, for convenience in summarizing data. This practice can

lead to distorted impressions of our data and inappropriate conclusions.

Thompson (1985) notes some alternative methods for analyzing data, including various OVA methods, regression coding, and commonality analysis. These methods each have their own advantages as well as disadvantages. The point is that there are many different ways to analyze data, depending on the type of data collected and the type of research involved. It is often appropriate to explore data in many ways.

Replicability

One final issue for consideration is that of replicability of one's results. The importance of this issue is noted by Carver (1978), who stated that " . . . replication is the cornerstone of science" (p. 392) and Thompson (1987), " . . . replication is the ultimate test of the generalizability of findings" (p. 32). Carver (1978) states that just because one's research was shown to be statistically significant does not mean the results are replicable. Shaver (1980) notes that statistical significance does not mean that treatment was necessarily effective.

Significance testing only alludes to whether errors of sampling are an alternative explanation of

results. In reality, there may be several other reasons which produced this result, such as internal validity threats. Researchers tend not to look past chance factors, not questioning how the research design and procedures themselves could have contributed to their findings. Design and procedures are not always carefully explained in studies, which has consequences for replicability.

As Thompson (1987) notes, suppose a Type I error was made in an original study which was thought to produce statistically significant results. A consequence of not carefully explaining design and procedures with regard to replication studies is that replication studies tend to only be published (or even submitted) if they are in support of the initial study. This means that instead of the replication studies sorting out the Type I error of the original study, the replication studies that are published all tend to support the original results which were associated with Type I error. Thompson (1987) has noted that no single study can give total support for causal hypotheses and that replications over time can sort out Type I errors for actual effects.

Conclusions

It seems that there is much more to analyzing and interpreting research results than just testing for statistical significance. Careful consideration must be given to other factors, especially the strength of the relationship between variables. As Thompson (1988) suggests, "Significance is not the end-all and be-all of research" (p. 18). Davidson and Giroir (1987) note "Researchers must accept . . . that significance testing is no longer the cornerstone for interpreting social science research results" (p. 10). A stronger view is held by Carver (1978) in his statement, "Apparently the case against such testing will have to be stated more loudly and more clearly to a wider audience if it is to have any effect" (p. 379).

Researchers should carefully consider what their research results really mean. Statistical procedures are too often used as substitutes to thought, rather than as aids to this process. It is time to return to some common sense base (Thompson, 1987). Shaver (1980) advocates teaching the theory behind the use of statistics, not just how to calculate. In doing so, thoughtful researchers will be produced and not persons

who foolishly rely on just statistical significance testing for interpretation of research results.

References

Carver, R. P. (1978). The case against statistical significance testing. Harvard Educational Review, 218(3), 378-399.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). New York: Academic Press.

Craig, J. R., Eison, C. L., & Metze, L. P. (1976). Significance tests and their interpretation: An example utilizing published research and ω^2. Bulletin of the Psychonomic Society, 7(3), 280-282.

Davidson, B. M., & Giroir, M. M. (1989, November). Comments on better uses of and alternatives to significance testing. Paper presented at the annual meeting of the Mid-South Educational Research Association, Little Rock, AR. (ERIC Document Reproductive Service No. ED 314 450)

Glass, G. V. (1979). Policy for the unpredictable (uncertainty research and policy). Educational Researcher, 8(9), 12-14.

Gold, D. (1970). A comment on 'a critique of tests of significance'. In D. E. Morrison & R. E. Henkel (Eds.), The significance test controversy--A reader (pp. 107-108). Chicago: Aldine Publishing Company.

Hays, W. L. (1988). Statistics (4th ed.). Ft. Worth: Holt, Rinehart and Winston, Inc.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1988). Applied statistics for the behavioral sciences (pp. 206-216). Boston: Houghton Mifflin Company.

Huberty, C. J. (1987). On statistical testing. Educational Researcher, 16(8), 4-9.

Keaster, R. D. (1988, November). Statistical significance testing: From routine to ritual. Paper presented at the annual meeting of the Mid-South Educational Research Association, Louisville, KY. (ERIC Document Reproduction Service No. ED 303 514)

Kupfersmid, J. (1988). Improving what is published. American Psychologist, 43(8), 635-642.

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. Journal of Consulting and Clinical Psychology, 46, 806-834.

Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. American Psychologist, 44(10), 1276-1284.

Rozenboom, W. W. (1970). The fallacy of the null hypothesis significance test. In. D. E. Morrison & R. E. Henkel (Eds.), The significance test controversy--A reader (pp.216-230). Chicago: Aldine Publishing Company.

Schneider, A. L., & Darcy, R. E. (1984). Policy implications of using significance tests in evaluation research. Evaluation Review, 8(4), 573-582.

Shaver, J. P. (1980). Readdressing the role of statistical tests of significance. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA. (ERIC Document Reproductive Service No. ED 191 904)

Thompson, B. (1985). Alternate methods for analyzing data from education experiments. Journal of Experimental Education, 54(1), 50-54.

Thompson, B. (1987, April). The use (and misuse) of statistical significance testing: Some recommendations for improved editorial policy and practice. Paper presented at the American Educational Research Association, Washington, DC. (ERIC Document Reproductive Service No. 287 868)

Thompson, B. (1988a, January). A note about significance testing. Measurement and Evaluation in Counseling and Development, 20(4), 146-148.

Thompson, B. (1988b, November). Common methodology mistakes in dissertations: Improving dissertation quality. Paper presented at the annual meeting of the Mid-South Educational Research Association, Louisville, KY. (ERIC Document Reproductive Service No. ED 301 595)

Thompson, B. (1989, March). Five steps for improving evaluation reports by using different data analysis methods. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA. (ERIC Document Reproductive Service No. ED 306 244)

Thompson, B. (1992). Two and one-half decades of leadership in measurement and evaluation. Journal of Counseling and Development, 70, 434-438.

Table 1

Statistical Significance at Various Sample Sized for a Fixed Effect Size (Moderate Effect Size)

Source	SOS	rsq	df	MS	Fcalc	Fcrit	Dec
SOSexp	98.7	0.098473	2	**49.35**	0.054614	200.00	NR
SOSunexp	903.6		1	903.6			
SOStot	1002.3		3	334.1			
SOSexp	98.7	0.098473	2	**49.35**	0.546148	4.10	NR
SOSunexp	903.6		10	90.36			
SOStot	1002.3		12	83.52			
SOSexp	98.7	0.098473	2	**49.35**	1.092297	3.49	NR
SOSunexp	903.6		20	45.18			
SOStot	1002.3		22	45.56			
SOSexp	98.7	0.098473	2	**49.35**	1.638446	3.32	NR
SOSunexp	903.6		30	30.12			
SOStot	1002.3		32	31.32			
SOSexp	98.7	0.098473	2	**49.35**	2.184594	3.23	NR
SOSunexp	903.6		40	22.59			
SOStot	1002.3		42	23.86			
SOSexp	98.7	0.098473	2	**49.35**	2.730743	c3.19	NR
SOSunexp	903.6		50	18.07			
SOStot	1002.3		52	19.27			
SOSexp	98.7	0.098473	2	**49.35**	3.276892	3.15	RJ
SOSunexp	903.6		60	15.06			
SOStot	1002.3		62	16.17			
SOSexp	98.7	0.098473	2	**49.35**	6.553784	3.07	RJ
SOSunexp	903.6		120	7.53			
SOStot	1002.3		122	8.22			

Note: Bolded numbers remain fixed. From Thompson (1989) with permission.

Table 2

Statistical Significance at Various Sample Sizes for a Fixed Effect Size (Larger Effect Size)

Source	SOS	rsq	df	MS	Fcalc	Fcrit	Dec
SOSexp	**337.2**	**0.336426**	**2**	**168.6**	0.253495	200.00	NR
SOSunexp	**665.1**		1	665.1			
SOStot	**1002.3**		3	334.1			
SOSexp	**337.2**	**0.336426**	**2**	**168.6**	2.534957	4.10	NR
SOSunexp	**665.1**		10	66.51			
SOStot	**1002.3**		12	83.52			
SOSexp	**337.2**	**0.336426**	**2**	**168.6**	5.069914	3.49	RJ
SOSunexp	**665.2**		20	33.25			
SOStot	**1002.3**		22	45.56			
SOSexp	**337.2**	**0.336426**	**2**	**168.6**	7.604871	3.32	RJ
SOSunexp	**665.2**		30	22.17			
SOStot	**1002.3**		32	31.32			
SOSexp	**337.2**	**0.336426**	**2**	**168.6**	10.13982	3.23	RJ
SOSunexp	**665.2**		40	16.63			
SOStot	**1002.3**		42	23.86			
SOSexp	**337.2**	**0.336426**	**2**	**168.6**	12.67478	c3.19	RJ
SOSunexp	**665.2**		50	13.30			
SOStot	**1002.3**		52	19.27			
SOSexp	**337.2**	**0.336426**	**2**	**168.6**	15.20974	3.15	RJ
SOSunexp	**665.2**		60	11.08			
SOStot	**1002.3**		62	16.17			
SOSexp	**337.2**	**0.336426**	**2**	**168.6**	30.41948	3.07	RJ
SOSunexp	**665.2**		120	5.542			
SOStot	**1002.3**		122	8.215			

Note: Bolded numbers remain fixed. From Thompson (1989) with permission.